Supporting Information

Raghavan et al. 10.1073/pnas.1205683109

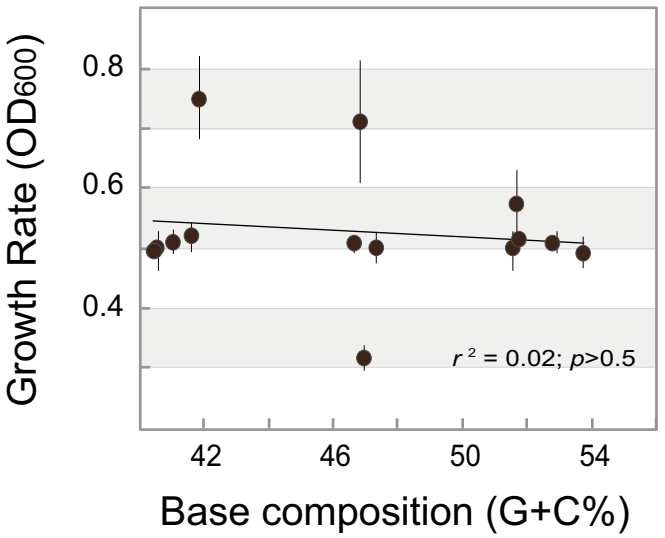

Fig. S1. The effect of G+C content on bacterial growth rates requires gene expression. Growth of *E. coli* strains, each containing a plasmid-borne GFP gene of a different base composition, was monitored hourly by measuring OD600 every hour. Values shown were taken at 5 h in strains that were not induced and represent the mean ± SD of three replicates.

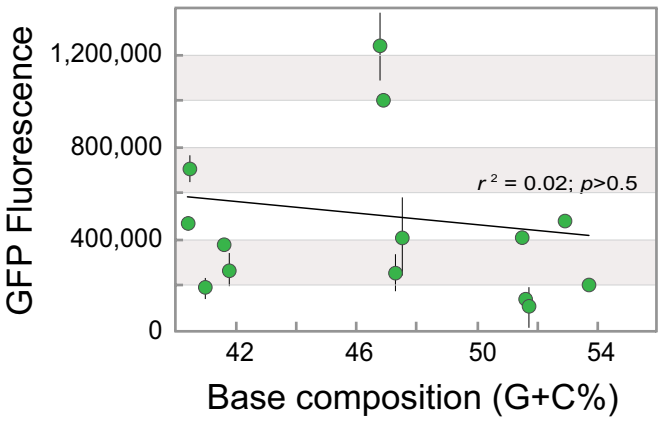

Fig. S2. No association between level of GFP expression and G+C contents of the GFP- gene. GFP fluorescence of *E. coli* strains expressing GFP genes that differ in G+C content was measured hourly. Shown are fluorescence values (*y* axis) corrected to OD600 measured at 5 h after induction. Values represent the mean ± SD of three replicates.

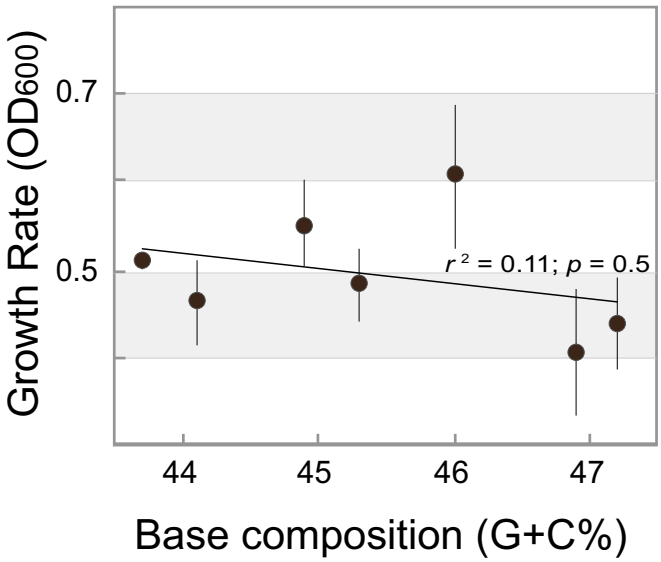

Fig. S3. Expression of *Bacillus* phage ϕ29 DNA polymerase genes is required to produce differences in growth rates. *E. coli* strains carrying plasmid-borne *Bacillus* phage ϕ29 DNA polymerase genes that differed in GC content from 43.7 to 47.2% grown without IPTG. OD600 values shown were taken at 5 h and represent the mean ± SD of three replicates.

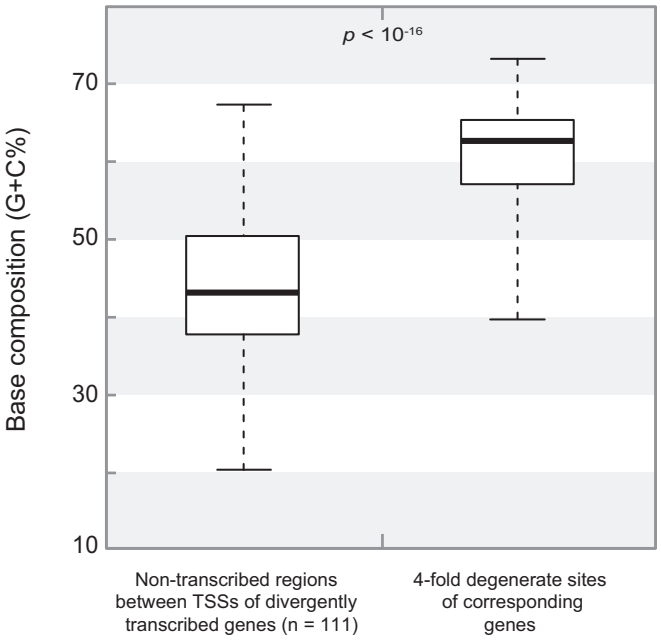

Fig. S4. G+C content of nontranscribed intergenic regions are significantly lower than that of fourfold degenerate sites in transcribed genes. The box plot shows the G+C% of fourfold degenerate sites for 111 pairs of adjacent, divergently transcribed genes (*Right*) and the G+C% of the nontranscribed regions between the transcription start sites (TSSs) of each gene pair (*Left*). Within each box, the heavy horizontal line marks the median G+C% value of the regions or genes considered, its top and bottom edges denote the first and third quartiles of the G+C% distribution, and whisker ends represent the range of G+C% values.

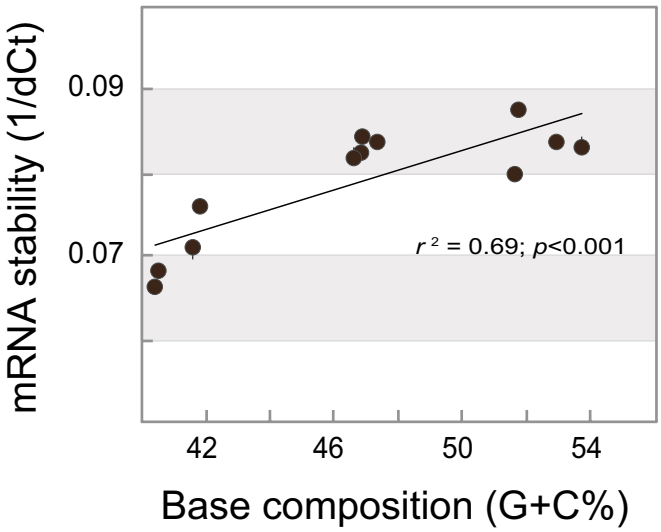

Fig. S5. Base composition correlates with mRNA stability. GFP mRNA stability was measured using qPCR after the induction of T7 lysozyme, for 15 min. The *x* axis shows the range of G+C % for the GFP genes. Inverse of dCt values (*y* axis) was used to calculate GFP mRNA prevalence and represent the mean ± SD of three experiments.

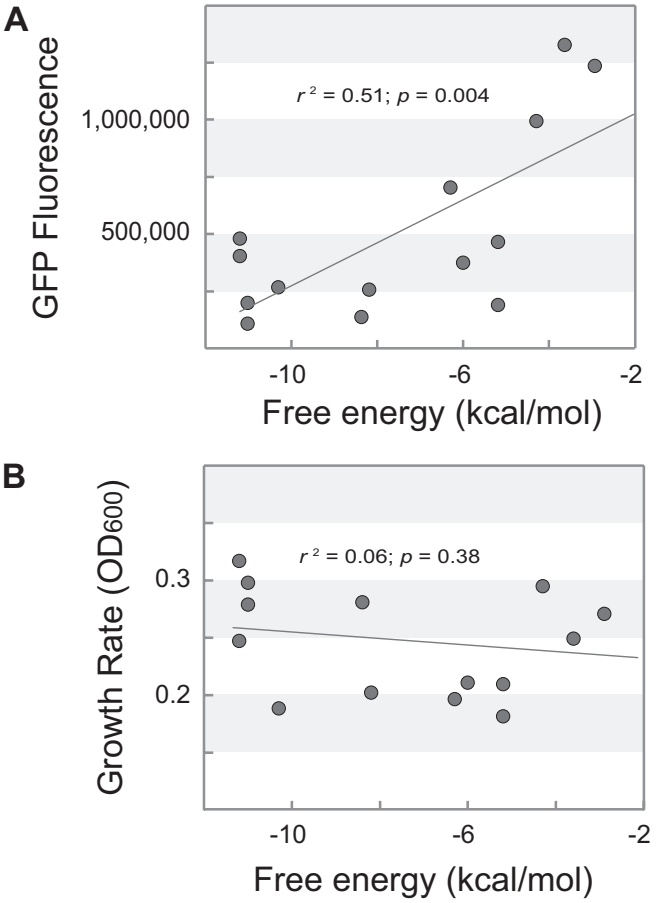

Fig. S6. Secondary structure of the 5′-end of mRNA is associated with protein production but not bacterial growth rate. Free energy (X-axes) was calculated for the 5′-end (-4 to +37 relative to translational start site) of mRNAs of GFP variants that differ in G+C content. Shown on Y-axes are GFP fluorescence values at 5 h after induction corrected to OD600 (*A*) and OD600 measured at 5 h after induction (*B*).

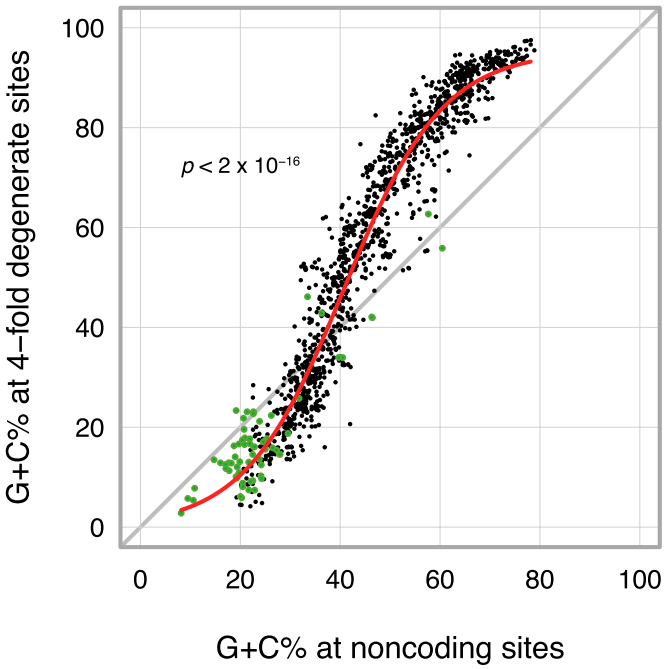

Fig. S7. Relationship between the average G+C content at fourfold degenerate sites of all genes except those encoding ribosomal proteins and the average G+C content of non- coding intergenic regions that are situated between convergently transcribed genes in fully sequenced bacterial genomes (n = 1430). Note that this plot, which removes biases that could arise from codon use preferences in highly expressed genes and from compositionally biased promoter regions, is virtually identical to that presented in Fig. 5, further confirming that the base compositions of allegedly 'neutral' fourfold degenerate sites are under different selective constraints than are noncoding regions. Green circles denote bacteria with genome sizes of less than one megabase (n = 67), and the red line indicates the logistic regression model fitted to the data.

Table S1. Detailed results of regressions in Fig. 1

Time (h)*	Slope	r^2	P
0	−0.0004	0.0260	0.5817
1	0.0016	0.1522	0.1679
2	0.0049	0.4002	0.0152
3	0.0062	0.4673	0.0070
4	0.0076	0.6715	0.0003
5	0.0078	0.7186	0.0001

*Hours postinduction. *E. coli* strains expressing GFP genes of different G+C contents were induced with 1 mM IPTG, and OD600 values were measured.